利用媒体大数据助力建设健康中国

——保健食品风险预警平台解构

摘　要：中国经济持续高速增长和人民健康意识的增强进一步带动了保健食品行业的快速发展。但保健食品行业乱象频出，引发了公众对保健食品的担忧，及对监管部门的不信任。中国健康传媒集团推出保健食品风险预警平台，协助监管部门规范保健食品行业，为公众健康保驾护航。

关键词：大数据；保健食品；风险预警
中图分类号：R-05　　文献标识码：A
文章编号：1671-0134（2019）02-046-03　　DOI：10.19483/j.cnki.11-4653/n.2019.02.005

文 / 王冲

随着经济高速发展，人民生活水平不断提高，健康意识也逐步增强，越来越注重养生保健。而保健食品作为具有特定保健功能的食品能够调节机体功能，越来越受到人民的青睐。2016 年 7 月 1 日正式施行的《保健食品注册与备案管理办法》将“保健食品”定义为：声称具有特定保健功能或者以补充维生素、矿物质为目的的食品，即适宜于特定人群食用，具有调节机体功能，不以治疗疾病为目的，并且对人体不产生任何急性、亚急性或者慢性危害的食品[1]。

2017 年，为全面落实《中华人民共和国食品安全法》和“四个最严”的要求，进一步加强对食品、保健食品生产、经营和进口单位履行主体责任的监管，严厉打击违规营销宣传产品功效、误导和欺骗消费者等违法行为，推动各地落实属地管理责任，强化企业守法诚信意识，营造健康有序的市场经营环境，切实保障消费者合法权益和消费安全，国务院食品安全办会同有关部门制定了《食品、保健食品欺诈和虚假宣传整治方案》。2019 年 1 月 8 日，国家市场监管总局、工信部、公安部、卫健委等 13 个部门召开电视电话会议，决定自 2019 年 1 月 8 日起，在全国范围内集中开展为期 100 天的联合整治“保健”市场乱象百日行动，该行动将在全国范围内加大对“保健”市场重点行业、重点领域、重点行为的事中事后监管力度，依法严厉打击虚假宣传、虚假广告、制售假冒伪劣产品等扰乱市场秩序、欺诈消费者等各类违法行为。

近年来，保健食品如雨后春笋般不断涌现，在给人民带来更多选择的同时，也出现了不少问题，如生产过程中违法添加药物、售卖假冒伪劣产品、夸大宣传产品效果等，这些隐患给人民造成了很大的困扰，也给市场监管带来了很大困难。为了让广大民众正确认识保健食品，同时协助监管部门加强对保健食品的管理，中国健康传媒集团率先推出保健食品风险预警平台。该平台是国内首家保健食品领域的风险预警平台，是以中文为核心的自然语言技术和保健食品业务应用紧密结合的知识成果；依托深度学习技术手段，结合计算语言技术与知识、社会计算技术与知识、人文计算技术与知识，通过样本训练、机器学习、人工经验专业判断相结合，进行系统、深入的研究，不断完善保健食品风险预警平台；该平台将通过业务合作、数据 API 服务等方式，为食品药品安全监管部门及相关机构提供数据与应用服务。

1. 中国保健食品风险预警平台

中国保健食品风险预警平台，内部叫法“保健食品冒烟指数”。“冒烟指数”构想来源于“森林开始冒烟是要发生火灾”，通过烟与火的形象比拟来推断冒烟指数与风险的关系，即从产品、企业冒烟指数高低来判断其风险的高低。

以“冒烟指数”为核心的风险预警平台大大降低了传统监管方式中运营成本的需求，与此同时，还提高了监管覆盖面和效率[2]。“保健食品冒烟指数”预警平台通过大数据采集、分析技术，全天候无缝隙地对保健食品进行监测预警。根据行业特点，从电商网站评论数据、贴吧、社交媒体等数据源大规模地采集原始数据，接入工商数据、法院执行数据等，并根据国产、进口、功能功效、适用人群等进行分类并存储在“保健食品冒烟指数”预警平台中，然后对多个数据源的数据，在内存式计算平台上进行分布式计算，经过数据清洗、数据集成等一系列预处理过程，把数据集合统一转换成可供分析的结构化数据。再通过分析算法，依照建立的多个风险预测模式，进行大数据的计算分析，用指数化方式从品牌，产品，价格等维度做风险排名（前 N 个）预警，在风险尚未酿成前及时发现，实现真正意义上的提前预警。

2. 本平台创新点

2.1 国内首家保健食品领域的风险预警平台，在平台构建、

关键词库建设等方面具有独创性

2.1.1 有利于及早发现保健食品风险，在风险尚未酿成前制止

通过分析研究保健食品活动的方式、特点以及演变规律，拓宽信息获取的渠道。监测平台以整合各地区、各有关部门信息资源，推动实现工商市场主体公示信息、互联网数据、法院立案判决执行信息等相关信息的互通共享，并抓取多种特征信息进行比对。利用监测平台的大数据技术，可以构建科学、专业的风险预测模型，通过风险预警模型计算出相应的风险指数，及早发现保健食品隐患，在风险尚未酿成前及时制止。

2.1.2 有利于及早打掉非法保健食品，在涉及面较小时遏制风险蔓延

当监测平台上企业风险程度高于设置的监控指数值时，系统会自动发出风险提示，需要监管部门密切关注。经过平台预警、对数据进行排查分析，及时做好相关的调查取证工作，快速进行登记，并按照部门职责分工，及时向相关部门进行交办。对于监测平台显示风险非常高的企业，在登记的同时及时向上级部门报告，监管部门及早介入，从而做到打早、打小，在涉及面较小时遏制风险蔓延，尽可能帮助公众挽回经济损失。

2.1.3 有利于节省风险排查耗费的人力资源成本

监测平台通过对汇总的各方数据（线下各部门的数据采集，线上的舆情公开数据，线上的监管企业数据上报），按照一定规则进行数据的清理整合，然后根据监测分析模型，根据已有的数据，计算出相关企业的风险指数及相关事件分析，最后根据计算出来的结果来进行相关处理（持续关注、人工干预、汇报情况等）。可以实时动态监测保健食品行业，相比通过传统的大量人工走访来摸排、电话文件传达等方式，可以节省很多风险排查人力资源成本。

2.1.4 有利于营造和谐环境维护社会稳定

对于违法违规保健食品企业，相关部门可以利用监测平台及早发现问题，缜密调查认定，迅速查办案件。另外，监测平台也能帮助法院和检察院加快案件审判。同时利用平台分析数据，可以针对性地对受害者开展宣传教育；利用定期发布的最新非法保健食品活动，不断提高公众防范非法保健食品的意识；剖析典型案例及其严重危害，降低非法保健食品的损害程度，维护社会稳定，营造和谐的社会环境。

2.2 大数据积累：准确、全面的收集数据是大数据的基础

2.2.1 数据库表设计

针对中国保健食品风险预警平台，对采集的企业数据信息进行合理设计数据库表；依据模型管理平台中模型所需的数据维度对数据表项进行分级，最大限度满足模型计算需要。

2.2.2 内部数据接入

数据采集平台支持 Excel 等格式的数据批量导入、通过接口与内部系统（总局数据、地方监管局数据）系统数据连通。

2.2.3 第三方数据接入

中国保健食品风险预警平台与美团、饿了么等数据提供商的接口进行连通。

2.2.4 对接工商数据

中国保健食品风险预警平台对目标范围内的企业工商信息进行对接，将企业的各类信息进行结构化存储入库，保证数据的质量和规模，入库信息数据准确度应达到 100% 与工商总局数据一致；制定数据更新规则，明确数据更新频率；对存量企业的数据定期更新。

2.2.5 采集电商数据

采集字段包括原文链接、商品名称、商品编号、商品价格、商品产地、规格（不同的规格对应不同的产品）、品牌企业、主要成分、商品毛重、适用人群、产品剂型、症状、蓝帽标识、国产 / 进口、保质期、京东自营、店铺、发货地、店铺名称、宣传图片、好评度、评论标签、全部评价数、晒图评论数、追评数、好评数、中评数、差评数等。对宣传图片进行采集，实现图片可追溯。

2.2.6 舆情系统

对接集团已有舆情系统，从指定网站或数据源，自动捕捉抽取相关产业、企业信息及政策信息，以链接或文字形式展示在相应区域中，能实现舆情的智能分析，包含情感分析、观点提取、地域分布、文章分类等，提供舆情可视化的监测平台。

2.3 自然语言处理：让机器更懂人类，提高关联性等特征分析

随着人工智能的大热，国内各大企业开始纷纷布局人工智能领域，并打造出各种不同的智能终端，比如人工机器人、无人驾驶汽车、智能电视、智能冰箱……这些智能终端有一个共同的特点——能读懂人类语言、与人类交流，还能进一步完成人类所下达的指令。

如此神奇的技术是如何实现的呢？这要归功于人工智能领域一项核心的处理技术——NLP。NLP（Natural Language Processing），即自然语言处理，它是研究人与计算机交互的语言问题的一门学科，也是人工智能一个重要的子领域[3]。简单来说，NLP 是让机器“理解”人们使用的自然语言结构和意思，将自然语言翻译为机器语言形式，并加工它（总结、句法分析等），再返回给用户自然语言。它涉及很多内容和技术，如文本朗读 / 语音合成、语音识别、中文自动分词、词性标注、句法分析、自然语言生成、文本分类、信息检索、信息抽取、文字校对、问答系统、机器翻译、自动摘要、文字蕴涵等。在人工智能发展之初，NLP 技术就已经显示出巨大的魅力。1949 年埃德蒙·伯克利（Edmund Berkeley）在他出版的《Giant Brains Or Machines That Think》一书

中写道：“最近出现许多消息，谈论的主题是奇怪的巨型机器处理信息，速度极快，技能很强……这种机器与大脑相似，由硬件和线缆组成，而不是血肉和神经，机器可以处理信息，可以计算、可以得出结论，可以选择，还可以根据信息执行合理操作。总之，这台机器可以思考。”[4] 作为人工智能核心技术之一，自然语言处理技术越发受到技术公司青睐，在国务院印发的《新一代人工智能发展规划》中，自然语言处理技术被列为关键共性技术。

2.4 模型构建具有独创性

2.4.1 风险监测分析挖掘模型设计

以监测对象为导向的风险监测分析挖掘模型——OFV。

风险分析 OFV 模型（监测对象 O– 风险特征 F– 特征值 / 特征程度 V），模型的构建由丰富的底层资源库和挖掘逻辑表达式支持。底层资源库包括监测对象库、业务知识分类库和监测要素网。风险特征是一系列特殊的文法生成式，其有限集合包含系统默认元素和资源库中各元素，系统默认元素（IR_）：内容、标题、时间、地区、关键词等；资源库中各要素：监测对象（O_）、业务知识（K_）和监测要素（E_）。

2.4.2 指数模型设计

指数模型综合考虑蓝帽标识、销量、负面评论数、星级指标、舆情传播数据等方面，训练语料获取最终值。

2.4.3 风险分级模型设计

根据风险计算得到的分值会划分为几个预警级别，其中正常监测为风险指数在 0–40 分的保健食品；重点监测为风险指数在 41–60 分的保健食品；高风险预警为风险指数在 61–80 分的保健食品；调查介入为风险指数在 81–100 分的保健食品。

结合各类数据与过往经验进行分析，利用数据挖掘、统计分析等技术方式建立保健食品风险研判模型，系统将按照最终形成的各类模型，在接入的海量数据中进行反复迭代和模式匹配，识别数据中存在的可疑行为，触发相应的预警，并建立风险指数体系，分析刻画保健食品的风险程度。

2.4.4 先进性

中国保健食品风险预警平台的建设，将风险把控关口前移，实现了从“事后舆情监测”向“事前风险预警”的跨越，为监管部门提供了重要的决策参考，对提升政府机构的公信力、促进产业健康发展、提升社会共治有重要意义。

本平台的建设，可以为有需要的党政机构、事业单位、企业提供大数据分析和数据的个性化定制等有偿服务；同时，树立集团大数据的品牌，为行业提供示范作用，提高品牌影响力；可以为领导掌握保健食品动态、研判形势、科学决策、指导工作提供有力的数据支持。

3. 平台建设意义

中国健康传媒集团是由国务院出资设立、财政部代为履行出资人职责的中央文化企业，行政管理和党的领导关系隶属于国家药品监督管理局。集团以服务食品药品监管、服务大众健康为宗旨，业务范围涵盖书、报、刊、网络（舆情监测与大数据）、教育培训、会展、影视、健康服务等健康传播领域。集团旗下目前拥有中国医药报社、中国医药科技出版社、中国健康传媒集团天创科技信息技术有限公司、中国健康传媒集团天和会展有限公司、天享（北京）健康科技发展有限公司、中国食品药品监管杂志社、中国食品药品网等多家新闻出版传媒文化单位。面向未来，中国健康传媒集团将致力于建设中国健康传媒领域的大型龙头企业，深度整合国家药品监督管理局所属的新闻宣传与传播出版资源，开拓更多适应市场需求的新型传媒业务，全方面、新视野诠释食品、药品、保健食品、化妆品、医疗器械及大健康产业新动态，为监管升级、产业发展以及大众健康服务。

中国健康传媒集团以其深厚的资源建设中国保健食品风险预警平台，防范保健食品“黑天鹅”事件，有利于监管部门及早发现保健食品行业潜在隐患，在规范保健食品行业方面发挥了重要作用。

保健食品监管是一项长期、复杂、艰巨的系统性工程，各级政府和各有关部门要高度重视；从保持经济平稳发展和维护社会和谐稳定的大局出发，切实增强防范和处置保健食品乱象工作的紧迫感、责任感和使命感，把思想和行动统一到中央决策部署上来，把防范和处置保健食品乱象作为一项事关全局的重要工作，进一步加大力度，最大限度地防范和化解风险，切实保护人民群众合法权益，坚决遏制保健食品乱象高发蔓延势头，坚决守住不发生区域性、系统性风险的底线，为经济社会持续健康发展提供有力保障。媒

参考文献

[1] 国家食品药品监督管理总局 . 保健食品注册与备案管理办法 [Z].2016–02–26.

[2] 李崇纲，许会泉 . 冒烟指数：大数据监测互联网金融风险 [J]. 大数据，2018，4（4）：76–84.

[3] 庞宇 . 基于中文 NLP 的计算机语言学关键问题的研究 [J]. 计算机产品与流通，2019（1）：236.

[4] Berkeley E C，Frankel S P.Giant Brains or Machines That Think[J].Physics Today，1950，3（5）：39.

（作者单位：中国健康传媒集团有限公司）